IJBST

ISSN: 0974 – 3987
IJBST (2011), 4(5):29-35

Feature Extraction using Balanced Multiwavelets for Classification of Microcalcification Clusters with application in breast cancer diagnostic

D.M.Garge[1], Dr.V.N.Bapat[2]

1 *Lecturer in Electronics, Government Polytechnic, Kolhapur, India, dattagarge@yahoo.com*
2 *Principal, College of Engineering, Miraj, India, vbkanhaji@gmail.com*

Abstract: This paper presents an approach for early breast cancer diagnostic by employing newer member of multiwavelet family: Balanced Multiwavelets. Detection and classification of microcalcification cluster is based on subband image decomposition. Detection of microcalcification is achieved by decomposing the mammogram and then reconstructing it from from the subbands containing only high frequency components. For this type of approach, we have applied different type of balanced multiwavelets for feature extraction. We used these results as an input to classification system. The proposed methodology is tested using Mammographic Image Analysis Society (MIAS) database. Results are presented as the receiver operating charactesistic (ROC) performance and quantified by the area under the ROC curve.

Key words: breast cancer, multiwavelets, balancing, microcalcification.

INTRODUCTION

Breast cancer is the type of cancer with the highest incidence rates in women. It is also the most common cause of cancer death in women in many countries like Asian countries and United States [1]. The early detection of breast cancer is vital to improve its prognosis. Moreover, it is well known that screening mammography is the best tool available for detection of cancerous lesions before clinical symptoms appear. Clustered microcalcifications are one of the earliest signs of potential cancerous changes in breast tissue. A microcalcification is a small calcium deposit that has accumulated in breast tissue, and it appears as small bright spot on the mammogram. Benign calcifications are generally larger, more rounded, smaller in number, more diffusely distributed, and more homogeneous in shape and size. Malignant calcifications are typically very numerous, clustered, small, dot like or elongated, variable in size and shape. However, because of small size of microcalcifications, the characterization of benign and malignant lesions represents a very complex problem even for expert radiologist.

In this work, we present a system for breast cancer diagnosis, and identification of microcalcification clusters in the digitized mammographic image and its classification as benign or malignant. To achieve the best results, we have employed different types of balanced multiwavelets such as BAT01, BAT02, CARDBAL2, and CARDBAL3. It is a well known fact that multi-scale representation has proven to be useful in many image processing applications and wavelet analysis is one way to generate such a representation. A newer alternative to wavelet analysis is the multiwavelet analysis which is very similar to wavelet analysis but has certain advantages those are proven to be more useful in image processing [2]. Because of the necessity of pre and post filtering of the image signal and its design constraints [3], researchers are giving attention to the balanced Multiwavelets. The concept of balancing was introduced in [4][5], extended to higher orders in [6][7], and further studied in [8]. In this paper, we have tried to elucidate the better performance of balanced multiwavelets in image processing applications such as feature extraction. We have attempted to use first, second, and third order balanced multiwavelets for feature extraction of mammographic images. This paper is organized as follows. Section II briefly presents the background survey of balanced multiwavelets. Methodology adopted and the algorithm implemented is covered in section III. Section IV focuses on results and related discussions.

BALANCED MULTIWAVELETS

It is known that multiwavelet filter bank is fundamentally a MIMO system. The problem of unbalanced channels used to split input signal is partially solved by adding some pre and post filters in the MIMO system. In pre-filtering, critical or non critical sampling is used which again destroys either the orthogonality or linear phase of the system or add redundancy. Hence, one can use orthogonal multiwavelet with good balance between the two scaling functions.

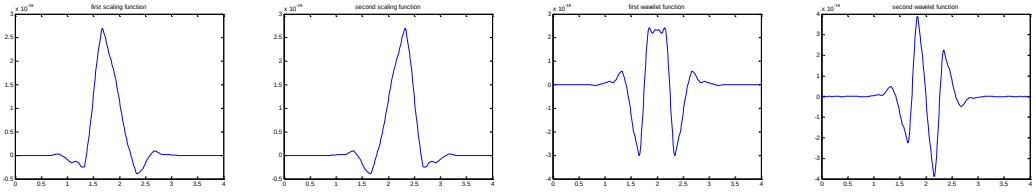
Balancing: In orthonormal case, defining the band-Toeplitz matrix corresponding to the low pass analysis,

$$L=\begin{bmatrix} \ddots & & & & \\ M[0] & M[1] & M[2] & M[3] & \cdots\cdots \\ & M[0] & M[1] & \cdots\cdots & \\ & & \cdots\cdots & & \\ & & & & \end{bmatrix} \quad (1)$$

where $\{M[k]\}_k$ is a sequence of r x r matrices of real coefficients. The condition imposed for balancing is that low pass synthesis operator L^T of the time varying filter bank based on this multiwavelet shall preserve the eigen signal [… 1, 1, 1, 1, …].i.e. L^T[… 1, 1, 1, 1, ……]^T = […… 1, 1, 1, 1, ……]^T. There are two ways to construct balanced multiwavelets. A simple way to construct balanced multiwavelets is to derive them from the complex Daubechies filters [9][10]. Another interesting way of constructing balanced multiwavelets is to balance already existing

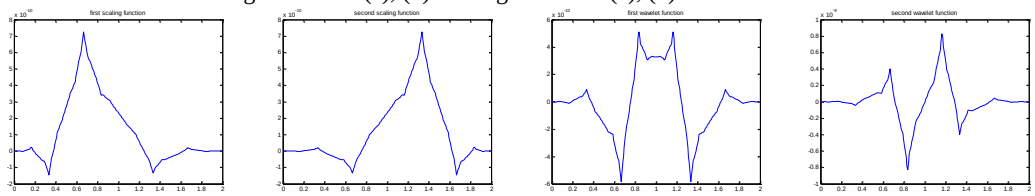
ISSN: 0974 – 3987
IJBST (2011), 4(5):29-35

unbalanced multiwavelets like the ones constructed in [11] [12]. Ultimately the aim of the balanced multiwavelet concept is to avoid intricate step of pre-filtering in multiwavelet based systems; that do not satisfy the interpolation / approximation properties of balancing. **Implementation of balanced multiwavelets:** We have implemented balanced multiwavelets (BMW) of compact support, with flipped scaling functions and symmetric / anti-symmetric wavelets for order one and two. These BMWs (BAT01, BAT02) are shown in fig. 1. The importance of balancing is also discussed by Selesnick [8] that the smoothness of scaling functions generated by h(n) and those by h(n-1) is entirely different. It is shown that GHM scaling function generated by h(n) are continuous and almost differential where as those generated by h(n-1) are extremely non-smooth as shown in fig. 2. The same behavior is evident for other unbalanced multiwavelets. In the same paper [8], it is also shown that the minimal length K-balanced multiwavelet bases based on even length symmetric FIR filters are better behaved than those based on odd-length symmetric FIR filters. For order 2 balancing, the minimal length of h_0 and h_1 are 8 and 12, supported on n=(0, . . . , 7) and n=(0, . . . , 11), respectively. Similarly order 3 balanced multiwavelets are of 12 and 16 minimal length. Fig. 3 show order 2 and 3 balanced multiwavelets.



(a) (b) (c) (d)

Fig. 1 BAT02 (a), (b) Scaling functions (c), (d) Multiwavelet



(e) (f) (g) (h)

Fig. 1 BAT01 (e), (f) Scaling functions (g), (h) Multiwavelet

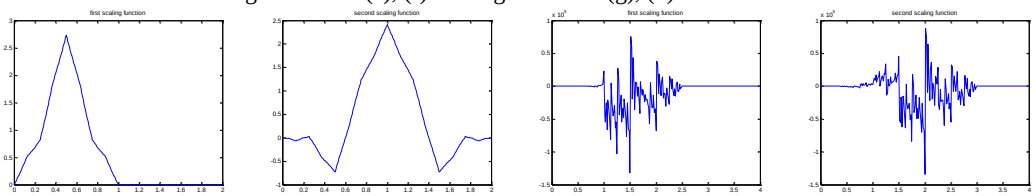
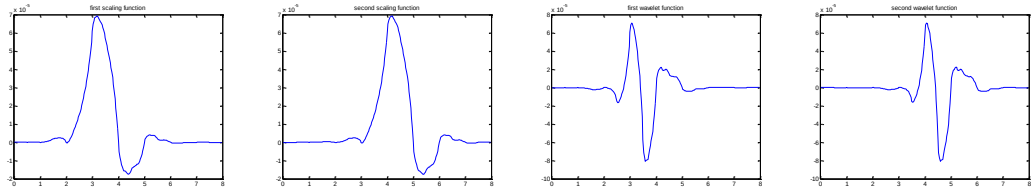
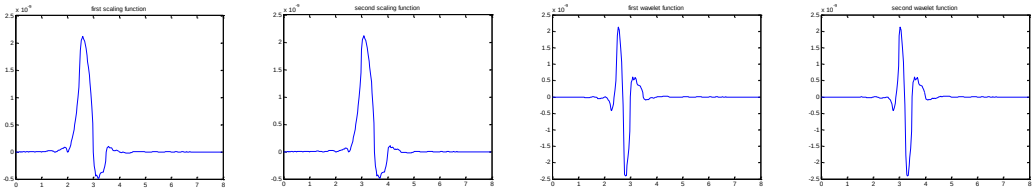


Fig. 2 GHM scaling functions generated by h(n) and h(n-1)



(a) (b) (c) (d)

Fig. 3 CARDBAL2 (a), (b) Scaling functions (c), (d) Multiwavelet



(e) (f) (g) (h)

Fig. 3 CARDBAL3 (e), (f) Scaling functions (g), (h) Multiwavelet

METHODOLOGY ADOPTED

Mammogram database is obtained from MIAS society developed by John Suckling [13] from which we have used 80 mammograms. Each image contains one or more clusters of microcalcifications verified by expert radiologists who have also marked the locations of

IJBST

ISSN: 0974 – 3987
IJBST (2011), 4(5):29-35

microcalcifications. Amongst 80 mammograms used for classification, 44 contain malignant lesions and 36 contain benign lesions.

Multiwavelet analysis is one way to generate multi-scale representation of the signal which is vital in image processing applications. Fig. 4 shows the final result of multiwavelet decomposition where first two columns and rows correspond to low pass filters (L_1, L_2), and the next columns and rows correspond to high filters (H_1, H_2). Thus we decomposed each image to 16 sub-images using each of the BMW. Fig. 5 shows 16 sub-images of a typical cluster of microcalcification decomposed to first level using CARDBAL3 BMW.
The method for feature extraction is based on decomposition of image and calculation of energy and entropy for image of each sub-band [14]. In an N × N sub-image, normalized energy and entropy are computed according to the following relations:

	L_1	L_2	H_1	H_2
L_1	L_1L_1	L_1L_2	L_1H_1	L_1H_2
L_2	L_2L_1	$L_2\,L_2$	L_2H_1	$L_2\,H_2$
H_1	H_1L_1	H_1L_2	H_1H_1	H_1H_2
H_2	H_2L_1	$H_2\,L_2$	H_2H_1	H_2H_2

Fig. 4 Image sub-bands after first level decomposition.

$$\text{Energy} = \frac{\sum_i \sum_j x^2{}_{ij}}{N^2} \quad (2)$$

$$\text{Entropy} = \frac{\sum_i \sum_j [x^2{}_{ij} / \text{Norm}^2] \log_2 [x^2{}_{ij} / \text{Norm}^2]}{\log_2 N^2} \quad (3)$$

where x_{ij} is the ij^{th} pixel value of the sub-image and, $\text{Norm}^2 = \sum_i \sum_j x^2{}_{ij}$

We generated 16 sub-images for every cluster using one level multiwavelet decomposition. As a result, we get 32 features for each microcalcification cluster.

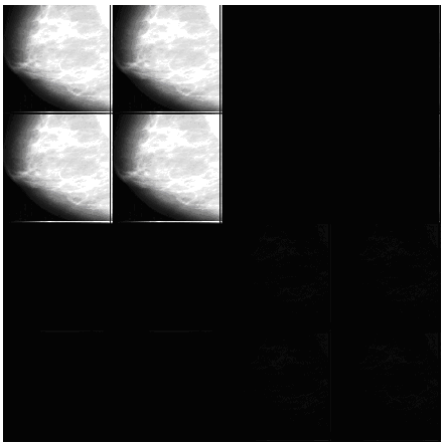


Fig. 5 Result after first level decomposition of mammogram image using balanced multiwavelet

Algorithm:

Detection of microcalcification cluster

A ⟵ Read mammogram image
[m n] ⟵ size of the image
For each pixel in A_{ij} {i=1, …,m; j=1, …,n}
Apply median filter with kernel of 3x3 size
End
C ⟵ Multiwavelet transform (A_{ij}) with BMW
[p q] ⟵ size of C
C[1:p/4, 1:q/4] ⟵ zeros (p/4, q/4); make low pass coefficients zero
D ⟵ Inverse multiwavelet transform (C_{pq})
th1 ⟵ threshold value for gray to binary conversion [16]
For each pixel in D_{ij} {i=1, …,p; j=1, …,q}
if (D_{ij} < th1)
D_{ij} = 0;
else
D_{ij} = 1;
end
D_{ij} ⟵ binary image showing microcalcification cluster

Classification of microcalcification cluster

A_{ij} ⟵ image after median filtering
B_{ij} ⟵ gradient filtering (A_{ij})
C ⟵ Multiwavelet transform (B_{ij}) with BMW
D_{xy} ⟵ subimage of C (x=1, 2, 3, 4; y=1, 2, 3, 4)
For each subimage D_{xy}
$Energy_{xy}$ ⟵ Energy as per Eq. 2
$Entropy_{xy}$ ⟵ Entropy as per Eq. 3
end
$Energy_{avg}$ ⟵ sum of $Energy_{xy}$ (x=1, 2, 3, 4; y=1, 2, 3, 4) / 16
$Entropy_{avg}$ ⟵ sum of $Entropy_{xy}$ (x=1, 2, 3, 4; y=1, 2, 3, 4) / 16
th1 ⟵ threshold value for energy feature [16]
th2 ⟵ threshold value for entropy feature
Classifier [16]
{
input ⟵ th1, th2, $Energy_{avg}$, and $Entropy_{xy}$
output ⟵ Benign / Malignant cluster
}

Fig. 6 psuecode of our system.

IJBST

ISSN: 0974 – 3987
IJBST (2011), 4(5):29-35

RESULTS AND DISCUSSIONS

As mentioned earlier, we have used balanced multiwavelets for extraction of features to be used by classification module. There are 32 features obtained from 16 sub-images.

Tables 1 to 4 show energy and entropy features where E11 corresponds to the energy or entropy value of the sub-image in the first row and first column of image shown in fig. 7

L_1L_1 (E11)	L_1L_2 (E12)	L_1H_1 (E13)	L_1H_2 (E14)
L_2L_1 (E21)	L_2L_2 (E22)	L_2H_1 (E23)	L_2H_2 (E24)
H_1L_1 (E31)	H_1L_2 (E32)	H_1H_1 (E33)	H_1H_2 (E34)
H_2L_1 (E41)	H_2L_2 (E42)	H_2H_1 (E43)	H_2H_2 (E44)

Fig. 7 Relation between decomposed sub-image and corresponding energy / entropy symbol.

Table 1 Energy data obtained using CARDBAL3 balanced multiwavelet applied to images.

	Mdb001	Mdb002	Mdb005-2	Mdb238	Mdb239-1	Mdb239-2	Mdb241
E11	46146.45	81260.27	13740.69	61921.46	69616.97	54673.50	27594.91
E12	45942.11	81208.58	13737.21	61932.34	69633.98	54596.91	27452.77
E21	46104.63	81195.57	13686.02	61906.02	69572.96	54602.53	27542.85
E22	45900.41	81145.98	13682.63	61916.92	69590.13	54525.99	27400.59
E13	1392.54	2369.04	404.23	1814.27	2038.05	1631.44	843.04
E14	1389.09	2463.29	406.81	1835.14	2064.71	1600.42	816.71
E23	1391.31	2366.92	402.65	1813.81	2036.67	1629.45	841.56
E24	1387.73	2460.86	405.09	1834.67	2063.50	1598.11	815.13
E31	1350.48	2367.79	392.90	1828.82	2065.41	1629.98	827.03
E32	1344.53	2365.98	392.79	1829.14	2065.83	1627.81	822.91
E41	1404.61	2471.29	452.19	1818.08	2038.08	1600.95	807.86
E42	1398.48	2469.33	452.06	1818.38	2038.66	1598.50	803.58
E33	40.98	69.03	11.57	53.59	60.52	48.60	25.25
E34	40.83	72.02	11.64	54.21	61.23	47.76	24.50
E43	42.54	72.26	13.29	53.28	59.64	47.81	24.72
E44	42.53	74.77	13.43	53.89	60.51	46.91	23.94

Table 2 Entropy data obtained using CARDBAL3 balanced multiwavelet applied to images.

	Mdb001	Mdb002	Mdb005-2	Mdb238	Mdb239-1	Mdb239-2	Mdb241
E11	0.9629	0.9918	0.9756	0.9988	0.9985	0.9964	0.9819
E12	0.9629	0.9920	0.9756	0.9988	0.9985	0.9965	0.9825
E21	0.9631	0.9920	0.9769	0.9988	0.9986	0.9965	0.9821
E22	0.9631	0.9921	0.9769	0.9988	0.9986	0.9966	0.9827
E13	0.9642	0.9912	0.9759	0.9986	0.9984	0.9962	0.9809
E14	0.9651	0.9892	0.9739	0.9986	0.9982	0.9957	0.9825
E23	0.9644	0.9913	0.9772	0.9986	0.9985	0.9963	0.9810
E24	0.9653	0.9894	0.9753	0.9986	0.9982	0.9959	0.9827
E31	0.9616	0.9914	0.9755	0.9988	0.9983	0.9962	0.9816
E32	0.9615	0.9916	0.9755	0.9988	0.9984	0.9963	0.9822
E41	0.9605	0.9882	0.9545	0.9986	0.9981	0.9958	0.9810
E42	0.9605	0.9884	0.9546	0.9986	0.9981	0.9959	0.9816
E33	0.9620	0.9907	0.9757	0.9986	0.9982	0.9960	0.9806
E34	0.9630	0.9883	0.9738	0.9985	0.9980	0.9955	0.9819
E43	0.9611	0.9871	0.9550	0.9984	0.9979	0.9955	0.9797
E44	0.9618	0.9857	0.9517	0.9983	0.9977	0.9951	0.9814

IJBST

ISSN: 0974 – 3987
IJBST (2011), 4(5):29-35

Table 3 Energy data obtained using CARDBAL3 balanced multiwavelet applied to images after gradient filtering.

	Mdb001	Mdb002	Mdb005-2	Mdb238	Mdb239-1	Mdb239-2	Mdb241
E11	91.30	607.19	956.71	2122.37	1638.64	2563.17	1748.60
E12	90.79	603.53	955.12	2122.11	1638.37	2562.83	1744.88
E21	83.74	576.23	734.51	1144.81	767.00	1618.27	818.87
E22	83.27	572.70	733.04	1144.51	766.76	1617.93	817.06
E13	3.70	18.57	28.46	62.57	48.46	75.59	52.80
E14	3.75	20.13	28.84	62.65	48.46	75.59	52.00
E23	3.38	17.62	21.88	33.80	22.78	47.78	24.99
E24	3.56	19.22	22.29	33.85	22.82	47.77	24.62
E31	23.20	408.58	73.08	266.25	352.43	251.23	331.30
E32	22.90	406.75	73.01	266.23	352.42	251.22	330.64
E41	11.45	136.53	79.01	272.06	243.97	280.04	261.62
E42	11.42	135.86	78.93	272.06	243.94	280.01	261.08
E33	1.15	12.05	2.18	7.85	10.40	7.41	9.95
E34	1.21	13.28	2.19	7.85	10.40	7.41	9.82
E43	0.85	4.11	2.36	8.02	7.21	8.26	7.87
E44	0.80	4.52	2.37	8.03	7.21	8.25	7.75

Table 4 Average of Energy data of table used for classification.

Mdb001	Mdb002	Mdb005-2	Mdb238	Mdb239-1	Mdb239-2	Mdb241
27.28	222.30	237.13	489.69	386.33	606.42	406.49

Table 5 Classification of clusters on five point scale (CARDBAL3 balanced multiwavelet)

Reference standard results	Radiologist's interpretation					
	Definitely benign	Probably benign	Possibly malignant	Probably malignant	Definitely malignant	Total
Benign	10	11	8	6	1	36
Malignant	2	3	6	20	13	44
Total	12	14	14	26	14	80

Table 6 Sensitivity, Specificity, and FPR for diagnosis and classification system of microcalcification clusters.

Test positive if greater than or equal to	Sensitivity	Specificity	False Positive Ratio (FPR)
2: Probably benign	0.955 (42/44)	0.278 (10/36)	0.722
3: Possibly malignant	0.886 (39/44)	0.583 (21/36)	0.417
4: Probably malignant	0.75 (33/44)	0.806 (29/36)	0.194
5: Definitely malignant	0.295 (13/44)	0.972 (35/36)	0.028

Note:- These data are obtained from the results in table. FPR is 1 minus specificity.

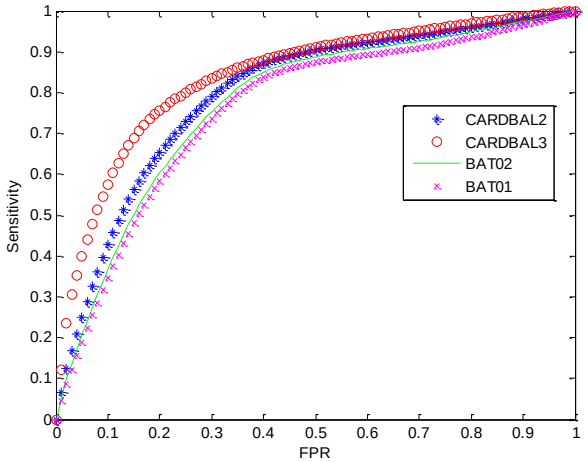


Fig. 8 ROC curves obtained from system corresponding to balanced multiwavelets.

(a) (b) (c)

Fig. 9 (a) Original Mammogram (b) Processed mammogram using BMW (c) Binary image

From Tables 1 to 4, it is evident that energy and entropy data can be used for classification of clusters. Microcalcification cluster images are obtained by selecting ROI from digitized mammogram images in which mdb001, mdb002, and mdb005-2 are benign lesions and mdb238, mdb239-1, mdb239-2, and mdb241 contain malignant lesions. Though energy and entropy data varies from image to image, the average of each image data is more suitable for classification, as seen from Table 4.

ROC analysis: Sensitivity and specificity, are defined as the number of true positive decisions / the number of actually positive cases and the number of true negative decisions / the number of actual negative cases, respectively. Sensitivity and specificity constitute the basic measures of performance of diagnostic tests [15]. Amongst 80 mammograms, 44 patients actually had malignancy and 36 patients had benign lesions; according to radiologist's interpretation given with the database. Lesions were interpreted according to five point scale: 1 (definitely benign), 2 (probably benign), 3 (possibly malignant), 4 (probably malignant), and 5 (definitely malignant). The result of this five point scale is shown in Table 5 and Table 6 shows four pairs of sensitivity / specificity values for ROC curve. Area under ROC curve is a combined measure of sensitivity and specificity. The ROC curves of our classification system are shown in the fig. 8 in which red colored curve is obtained using CARDBAL3, blue colored curve from CARDBAL2, green colored from BAT02, and magenta colored from BAT01 balanced multiwavelets. The area under ROC curve is computed which indicates that the CARDBAL3 balanced multiwavelet has proven to be a better choice for classification of microcalcification clusters.

Conclusion

Intelligent system has been developed for the identification and classification of microcalcification clusters in digitized mammograms. This system is meant as a second opinion to assist radiologists for better interpretation of mammograms and to enhance the diagnosis reliability. The method employs balanced multiwavelets for experimentation on MIAS database of mammograms. As seen from the area under ROC curve, CARDBAL3 balanced multiwavelet with area ranging around 0.8398, outperforms the other BMWs with areas CARDBAL2 (0.8007), BAT02 (0.7757), and BAT01 (0.762). The system successfully combines intelligent methods and image processing techniques to enhance mammographic diagnosis sensitivity (Fig. 9). The proposed methodology could be a part of an integrated CAD system which could assist radiologists in mammogram analysis and diagnosis decision making.

REFERENCES

[1] CancerNet, A service of the National Cancer Institute, http://cancernet.nci.nih.gov

[2] M. Vetterli and G. strang (1994) Time varying filter banks and multiwavelets. in Proc. 6th IEEE Digital Signal Processing Workshop, Yosemite, CA.

[3] X. G. Xia, J. S. Geronimo, D. P. Hardin, and B. W. Suter (1996) Design of pre-filters for discrete multiwavelet transforms. IEEE Transaction Signal Processing, 44: 25-35

[4] J. Lebrun and M. Vetterli (1997) Balanced multiwavelets. in Proc. IEEE ICASSP, Munich, Germany: 2473-2476

[5] J. Lebrun and M. Vetterli, (1998) Balanced multiwavelets theory and design. IEEE Transactions on Signal Processing, 46(4): 1119-1125

[6] J. Lebrun and M. Vetterli, (1998) High order balanced multiwavelets. in Proc. IEEE ICASSP, Seattle, WA: 1529-1532

[7] J. Lebrun and M. Vetterli, (2001) High order balanced multiwavelets: theory, factorization, and design. IEEE Transactions on Signal Processing, 49(9): 1918-1930

[8] I. W. Selesnick (2000) Balanced multiwavelet bases based on symmetric FIR filters. IEEE Transactions on Signal Processing, 48(1): 184-191

[9] W. Lawton (1993) Applications of complex valued wavelet transforms to sub-band decomposition. IEEE Transactions on Signal Processing, 41: 3566-3568

ISSN: 0974 – 3987
IJBST (2011), 4(5):29-35

[10] J. M. Lina and M. Mayrand(1995) Complex Daubechies wavelets. Appl. Comput. Harmonic Anal., 2: 219-229

[11] C. K. Chui and J. A. Lian (1996) A study of ortho-normal multiwavelets. Appl. Numer. Math., 20(3): 273-298

[12] J. S. Geronimo, D. P. Hardin, and P. R. Massopust (1994) Fractal functions and wavelet expansions based on several scaling functions. J. Approx. Theory, 78(3)

[13] Mammographic Image Analysis Society MiniMammographic Database, http://skye.icr.ac.uk/miasdb/miasdb.html

[14] H. Soltanian-Zadeh, F. Rafiee-Rad, and S. Pourabdollah-Nejad (2004) Comparison of multiwavelet, wavelet, Haralick, and shape features for microcalcification classification in mammograms. Pattern Recognition, 37: 1973-1986

[15] S. H. Park, J. M. Goo,and Chan-Hee Jo(2004) Receiver operating characteristic (ROC) curve: practical review for radiologists. Korean J. Radiol, 5(1): 11-18

[16] D.L.Donoho, "Nonlinear wavelet methods for recovery of signals, densities, and spectra from indirect and noisy data", Proc. Symposia Applied Math, Vol. 47(1), pp. 173-205 (1993).